(* COUNTS *)

25

Beginning to Display Counts

(patterns of distribution)

		Exh.	Page
Introduction	25-1		
		a	
		b	
		1	25-8
		2	25-9
		3	25-11
25A.	Some matters of principle 25-4		
25B.	Variations on the stem-and-leaf theme 25-7		
		c	
		4	25-13
		5	25-14
		6	25-16
		7	25-18
25C.	Variations on the histogram 25-10		
25D.	Intervals of different lengths 25-19		
	* * *		
		d	
		8	25-21
		9	25-22
		10	25-24
		11	25-25
		12	25-27
25H	How far have we come? 25-28		
	* * *		
25K.	Xmas-tree patterns (background) 25-29		
	* * *		
		e	
		k	
		13	25-30
		14	25-31
		15	25-32
25P.	Problems for chapter 25 25-33		
		p	
		16	25-34
		17	25-36
		18	25-37
		19	25-39
		20	25-40
		21	25-41

25

Beginning to Display Counts
(patterns of distribution)

In chapter 1 we met stem-and-leaf displays. For small or moderate batches of values, they gave us a reasonable and useful picture of where the values were located and how they were spread out. As batches grow larger, we are less and less satisfied with stem-and-leaf displays, though we have to note that they have no effective competitor if we want to, at the same time, both give a general idea and provide relatively full detail. Schematic plots give less detail.

So far as giving a graphic general idea about a batch of numbers goes, all the methods of high usefulness that we know depend on counts -- on counts of the number of observations falling in selected intervals. This is so whether the basic numbers are amounts -- or counts -- or balances -- or re-expressed forms of amounts or counts. The first and main aim of this chapter is to explore these methods.

Once we can deal effectively with counts of numbers, we naturally seek to learn how to deal with other situations. If the categories into which the observations fall are ordered, but not numerical, we do have a structure in whose terms we can look at counts. The techniques that prove to be natural and effective in dealing with ordered counts are somewhat different than those we use for counts of sizes, but, as we

will see in chapter 28, the underlying philosophy is much more nearly the same than may appear at first sight.

What if the things that we are counting do not even have a natural order? Occurrences of words are a case in point. Words do not, in fact, have a meaningful order. Dictionary order is an essential convenience for many purposes. For anyone interested in the use of words as a whole, however, information about the number of times each word was used in a body of text or speech is of near-minimum usefulness if it is given for words arranged in dictionary order. When we have nothing else, we use what we can. If we shift our attention from the names of the words to the number of times they appear, we can make reasonably helpful displays. We now count the number of words which appear a given number of times. In simple language, we count the counts. Looking at counts of counts can reasonably be done in varied ways, as we will see in chapter 27.

How are a set of observed numbers distributed? Any answer to this question, however sophisticated its mechanism, is likely to begin with counts -- counts of how many cases fall in each of certain intervals or counts of how many cases take on each of the possible values. The devices that we use to display apparent behavior -- or to compare observed behavior to a fitted partial description -- need to be devices appropriate to counted data. Now that we have a little background in dealing with counts, we are ready to tackle the question of exhibiting how counts are distributed among classes -- of

what counted distributions look like -- and how our looks at them can be made more searching and more effective. We will need a new principle or two, added to those that are now becoming familiar, but we will find that looking at counted distributions again reveals many of the same aspects and approaches as have our other topics.

In doing this we have to recognize the differences among different kinds of classifications:

- classes that have names but are not ordered according to some definition of <.

The countries of Europe or the countries of sub-Saharan Africa would be examples of this.

- classes that are arranged in order according to some <.

These may be classes that exist in themselves or classes defined for our convenience. Justices of the Supreme Court rank above senators, who rank above congressmen, the boundaries being widely recognized and sharp. Colonels, majors, captains and lieutenants would provide another example of classes with independent existence. Classes of people aged 20-24, 25-29, 30-34, 35-39, and so on, offer a simple example of classes defined for our convenience.

In most instances where ordered classes exist for our convenience, there is an underlying numerical scale. This case will be the focus of our interest in this chapter.

25A. Some matters of principle

We already have some background that is unfamiliar to many; we have been using stem-and-leaf arrays to display distributional information. As one particular result, we are used to seeing pictures of patterns of values in which the different values are spread out vertically.

It cannot be a matter of too high principle whether pictures of distributions spread the values out horizontally or vertically. Clearly both can serve us. But it is important for us to understand why sometimes one choice is made and sometimes another.

Whether the usual division between what is vertical and what is horizontal in a graph is intrinsic in Homo sapiens or is an accident of Western civilization is a difficult question, one that is hard to even see how to attack. Fortunately, we do not need to know its answer. We live in a world in which responses are plotted up and down, while quantities that describe alternatives -- whether or not it is likely that they "affect" the responses -- are plotted from left to right. (Is the fact that North is at the top of our maps a consequence of the West-East trend of the Mediterranean Sea?)

There are two quite disparate views about a plot of a distribution of numerical values. One view thinks -- often only implicitly -- about the numbers of cases (per interval or per possible value) as what matters. It usually thinks of these numbers as responses and about the values involved as

something that influences or describes the environment of such counts. Those who -- for the moment at least -- think in this way naturally spread values out horizontally and plot counts vertically. (Within our tradition we could expect nothing else.)

The other approach, typified by the stem-and-leaf array, thinks -- again often only implicitly -- of the values as what matter -- as what ought to be displayed clearly. Those of this mind will spread values out vertically, and use some graphical device to make clear the frequency with which various ranges of values occur.

Each approach has its place. We should all try to use each in its place.

How do we picture importance most directly? How could we convey relative frequencies visually to a preliterate -- be he uneducated Australian bushman or preschool child? We leave the final answer to this question to the psychologists, but give a simple approximate answer -- pebbles.

To show one thing 100 times as important -- or 100 times as frequent -- as another we could use 100 pebbles for one but one pebble for the other. To make clear the comparison we would be very likely to lay out the 100 pebbles side by side -- but probably not in a line. A square 10x10 array of pebbles will make the point, and be easy to comprehend. (In the strict word-genealogy sense of "comprehend" -- grasping together.) All of us have played with pebbles and blocks and

many similar things, whether we remember this or not. Almost all of us have developed at least somewhat of a feel for area -- for area as representing importance or count. As a result, whenever we can, without great loss of some other sort, use a scheme of display in which areas are proportional to what interests us -- here counts -- we ought to do so.

The problems of effective representation of counted distributions, then, revolve around how to combine as much as possible of:

- proper use of vertical or horizontal as the direction of spreading out values. (The cultural principle.)
- sticking to area proportional to count. (The area principle.)
- gaining graphical effectiveness by making the (implicit) standard of comparison a straight line, preferably a horizontal (or vertical) one. (The comparison standard principle.)
- seeking out the advantages of making residuals evident and easily scrutinized. (The residuals principle.)
- following out our experience which favors expressing counts by their square roots -- simple or double. (The square-root principle.) [We have seen advantages, in some cases, of going beyond square roots to logarithms. They do not seem to be common in the present situation.]

Two of these principles are new to us, one comes from chapter 24, the other two are old. To combine all five in a single picture often calls for real ingenuity.

25B. Variations on the stem-and-leaf theme

Stem-and-leaf displays meet our two new criteria: Values are spread out vertically. Using one digit per case means that the area covered with digits is proportional to count -- at least on an ordinary typewriter. Thus it will help us to prepare for more complex cases if we look at a few variations on the skeleton of the stem-and-leaf structure.

Exhibit 6 of chapter 17 (page 17-10) shows 120 second residuals from a Nebraska voting example. It does this in what we now consider a conventional stem-and-leaf pattern. Exhibit 1 repeats this data in the same basic stem-and-leaf pattern, swapping up with down so as to leave positive values higher than negative ones.

So long as we think of a stem-and-leaf display as something printed -- as a conventional table or something produced line by line by a computer printer -- it is reasonable for it to begin with the most negative values and move toward the most positive ones. If we are to think of a picture, however, the most positive values should be at the top.

The simplest way to give the same information purely graphically -- we must stress purely because any stem-and-leaf display is largely graphical -- is to use "bars flying from a mast", in which the length of each bar is proportional to the corresponding count. Exhibit 2 shows the result. Except for the elimination of some distraction, this has very little to offer beyond exhibit 1. If we were dealing with large

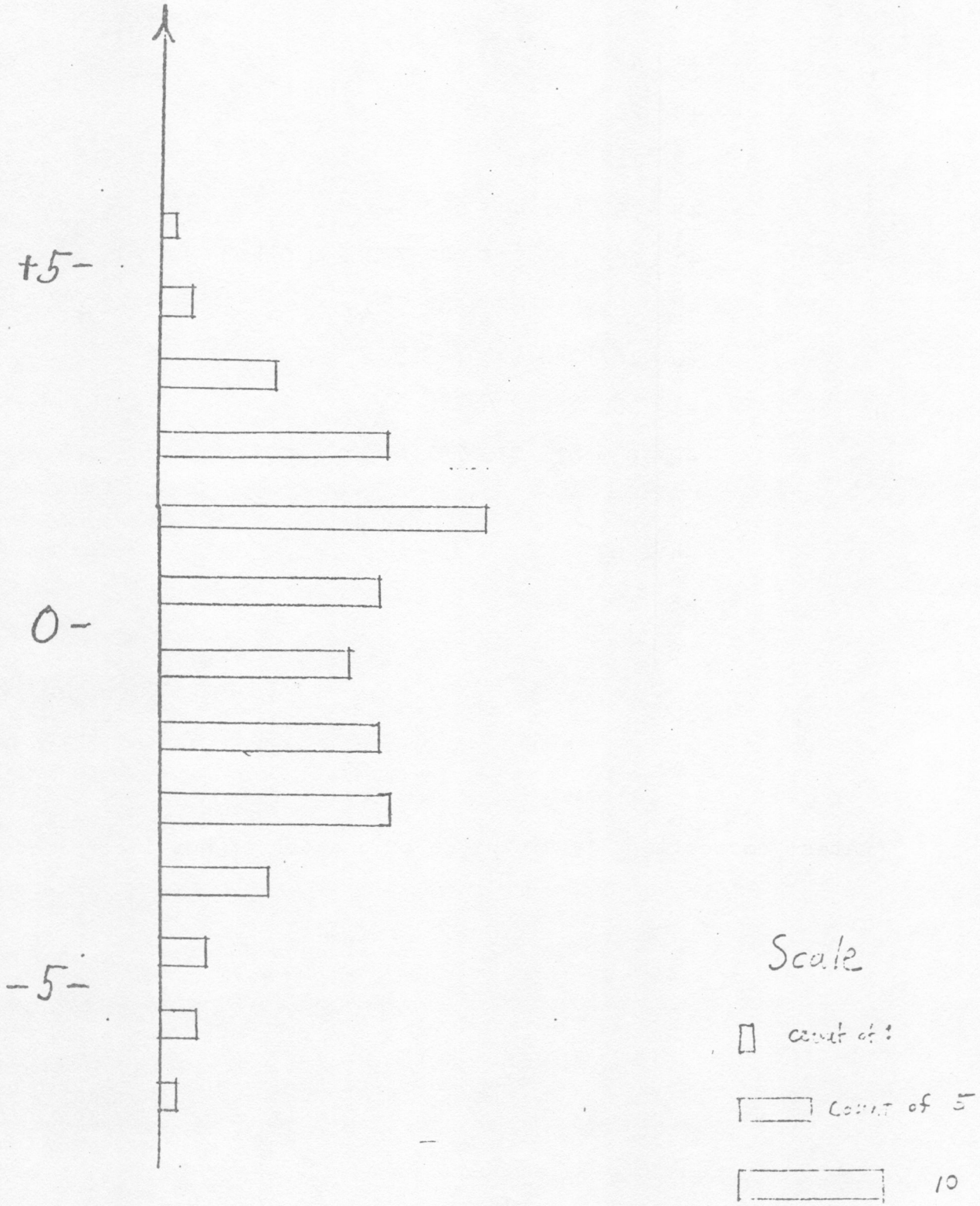
exhibit 1 of chapter 25 -- Nebraska residuals
Plus-side-up version of exhibit 6 of chapter 17
(120 residuals from an elementary analysis
of a two-way table)

+5	4
+4	22
+3	118 339 26
+2	702 431 448 633 842 0
+1	263 133 347 949 792 250 014 3
+0	181 413 624 267 768
-0	167 048 443 743 5
-1	266 128 306 576 616
-2	930 016 070 791 645 6
-3	452 212
-4	203
-5	17
-6	1

(Latest corrections to exhibit 6 of chapter 17 not incorporated.)

exhibit 2 of chapter 25 -- Nebraska residuals

The 120 residuals Shown by "bars flying from a mast"
(compare exhibit 1)



counts, where direct use of the stem-and-leaf display would be impossible or confusing, the bars would have a real advantage.

We ought next to see what we can do to give effect to chapter 24's view that counts are most usually well expressed numerically by their square roots (or by a logarithmic expression). At first sight, such a square-root principle seems squarely in contradiction with the area principle, which calls for equal areas for equal counts. But this need not be so. Bars have both height and length. If we consider bars both of whose dimensions increase as the count increases, while retaining the same proportion, we can have the linear dimensions of our bars proportional to the square root of the count, while their area is proportional to the count itself.

Exhibit 3 shows the same distribution of values described by such bars, bars which are naturally called (square) rooted bars, or, more clearly

rootobars

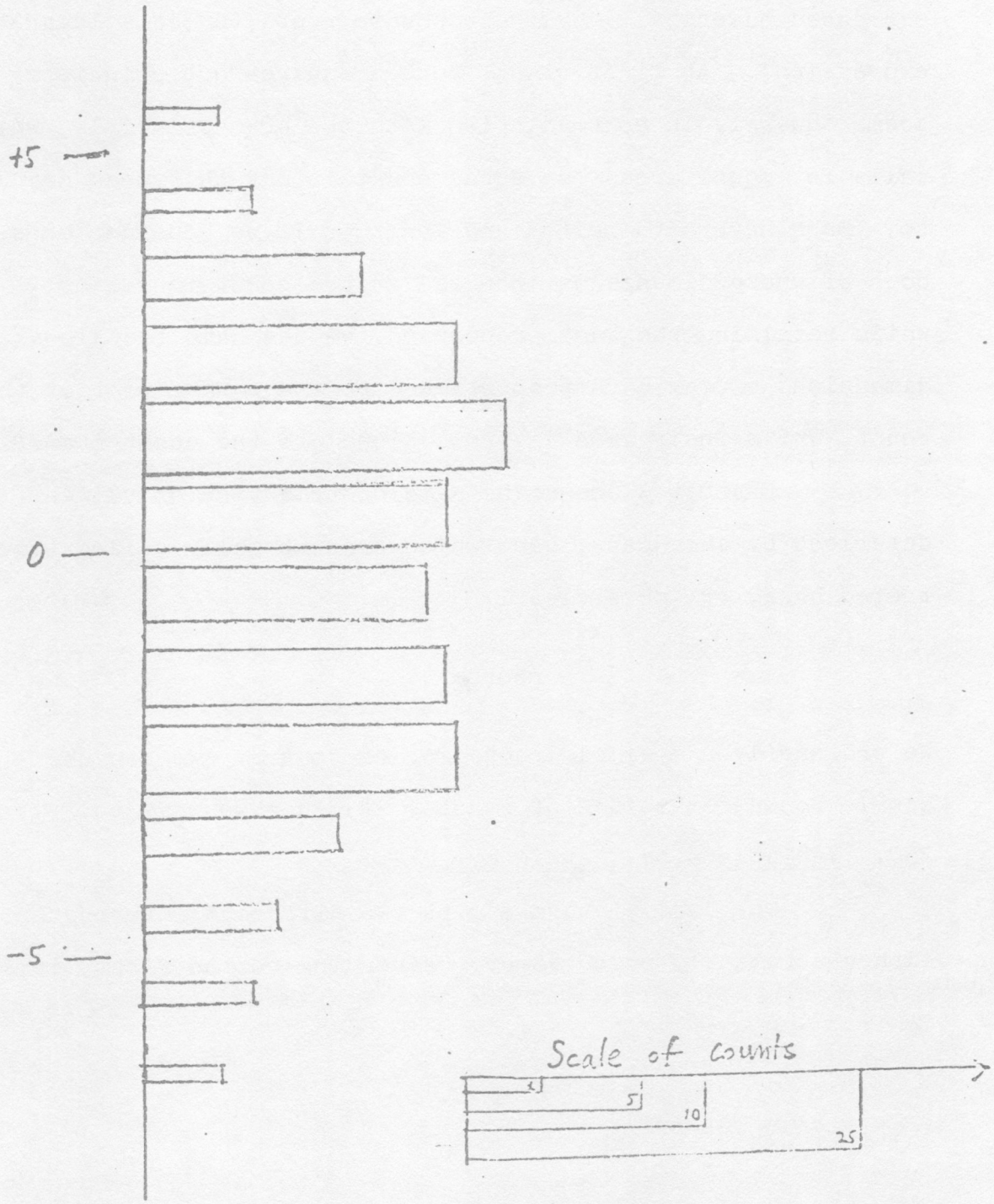
We ordinarily use simple roots for rootobars. We can use double roots (exhibit 2 of chapter 24) when we want to.

25C. Variations on the histogram

Our last picture admits two different interpretations -- either separately or together. According to the first, it is

exhibit 3 of chapter 25 -- Nebraska residuals

The same 120 residuals, Shown by rootobars "flying from a mast"



a display of the distribution of the values of 120 residuals, with equal area given to each value. As such, it is rightly plotted with the values (of the residuals) running down to up.

According to the second interpretation, it shows -- for each interval of length one that can be defined by giving the units digit -- the count associated with that interval, this count being expressed on a square root scale. To the extent that this latter interpretation is correct, our picture would be more appropriately plotted with values running from left to right, thus reserving down to up for count information which is now clearly regarded as a response.

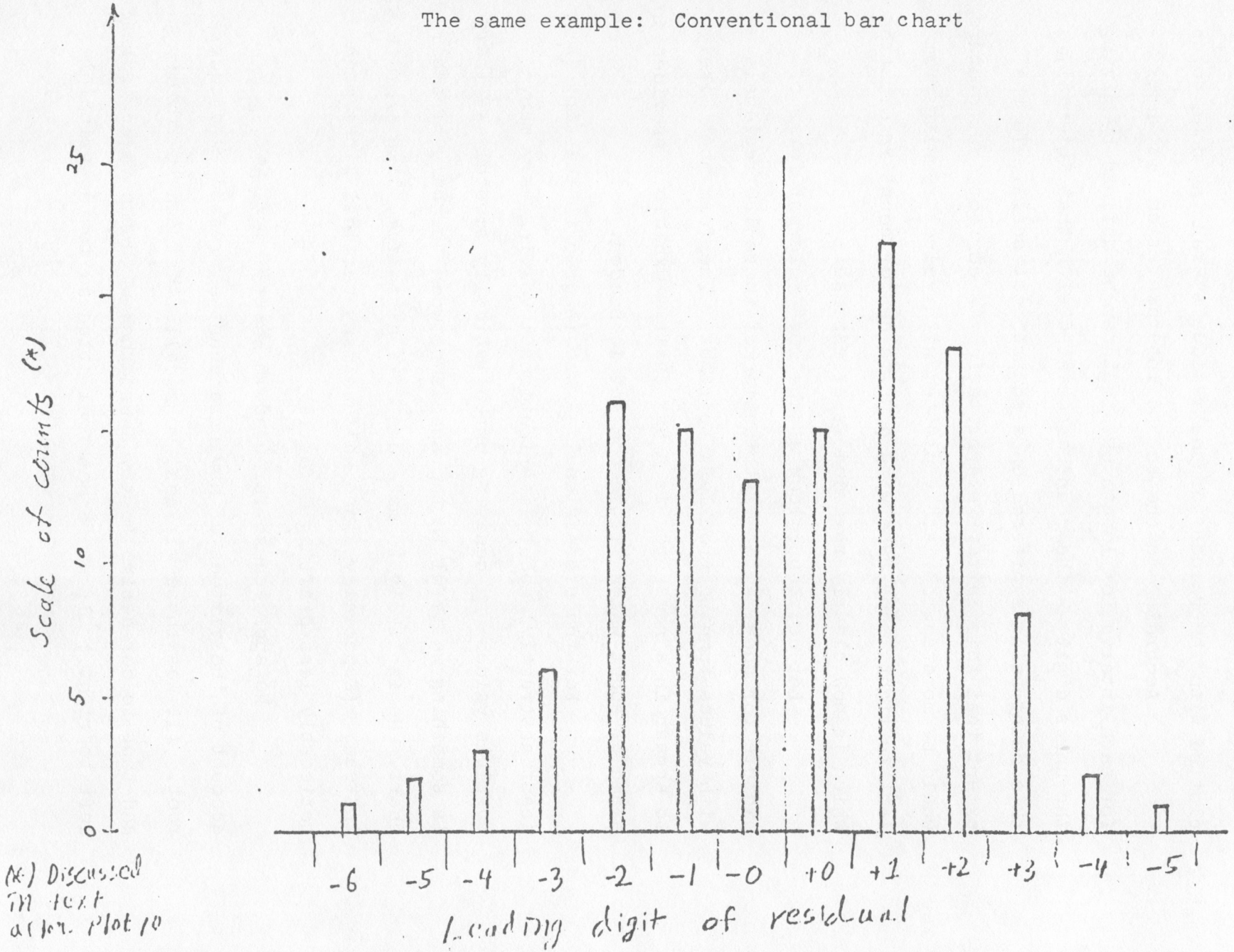
Let us, then, look at the most relevant forms of picture for showing the distribution of these 120 residuals with values running from left to right. The form most likely to appear in a newspaper is a bar chart. Exhibit 4 provides an example, about which little comment is needed.

For the professional -- who is used to thinking (at least abstractly) about distributions involving very many possible values and behaving unrealistically smoothly -- there is something wasteful in the spaces between the bars of a bar chart. He is likely to make use of a <u>histogram</u>, like exhibit 5, instead. (Especially since half as many vertical lines mean noticeably less drafting effort.)

No approach so far used in section 25B takes any account of the virtues of having a count expressed as a square root. (It was these virtues, we recall, that led to exhibit 3 -- and thus to our desire to try approaches with the values spread left to right.) If we do not feel attached to the principle

exhibit 4 of chapter 25 -- Nebraska residuals

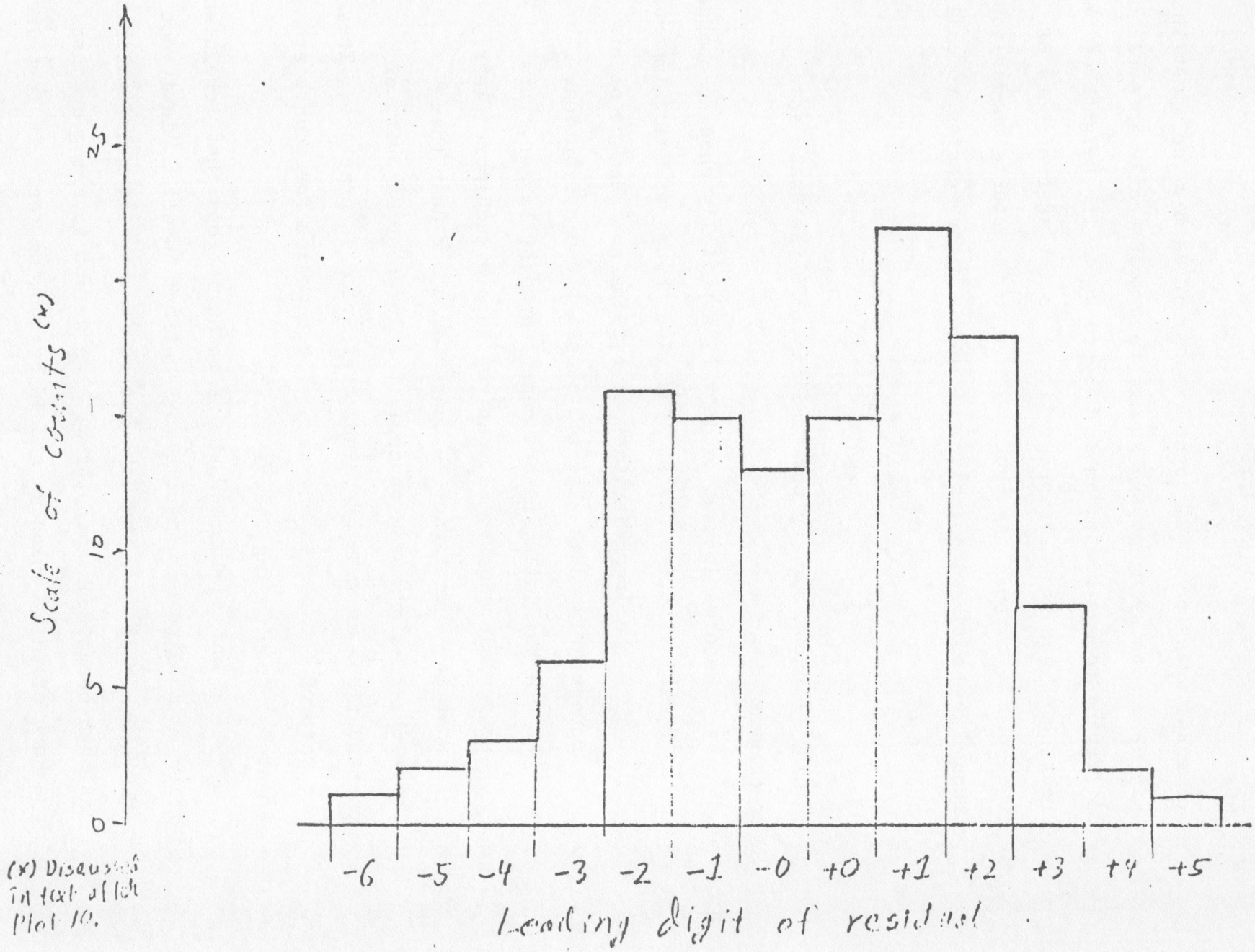
The same example: Conventional bar chart



(*) Discussed in text after Plot 10

ıibit 5 of chapter 25 -- Nebraska residu

The by-now-familiar example: Histogram



(x) Discussed in text after Plot 10.

of equal area for equal count -- some of us may feel just this way -- we can bring in the square root most easily in a picture differing from a histogram in but one way: The heights of the rectangles are now made proportional to the square roots of the counts, rather than proportional to the counts themselves. Though probably bad philology, it makes memory easier to call such a plot a

rootogram

Exhibit 6 shows our, by now familiar, 120 residuals as a rootogram.

When we look at exhibit 6, especially when we compare it with exhibit 5, we see at least two things of major importance:

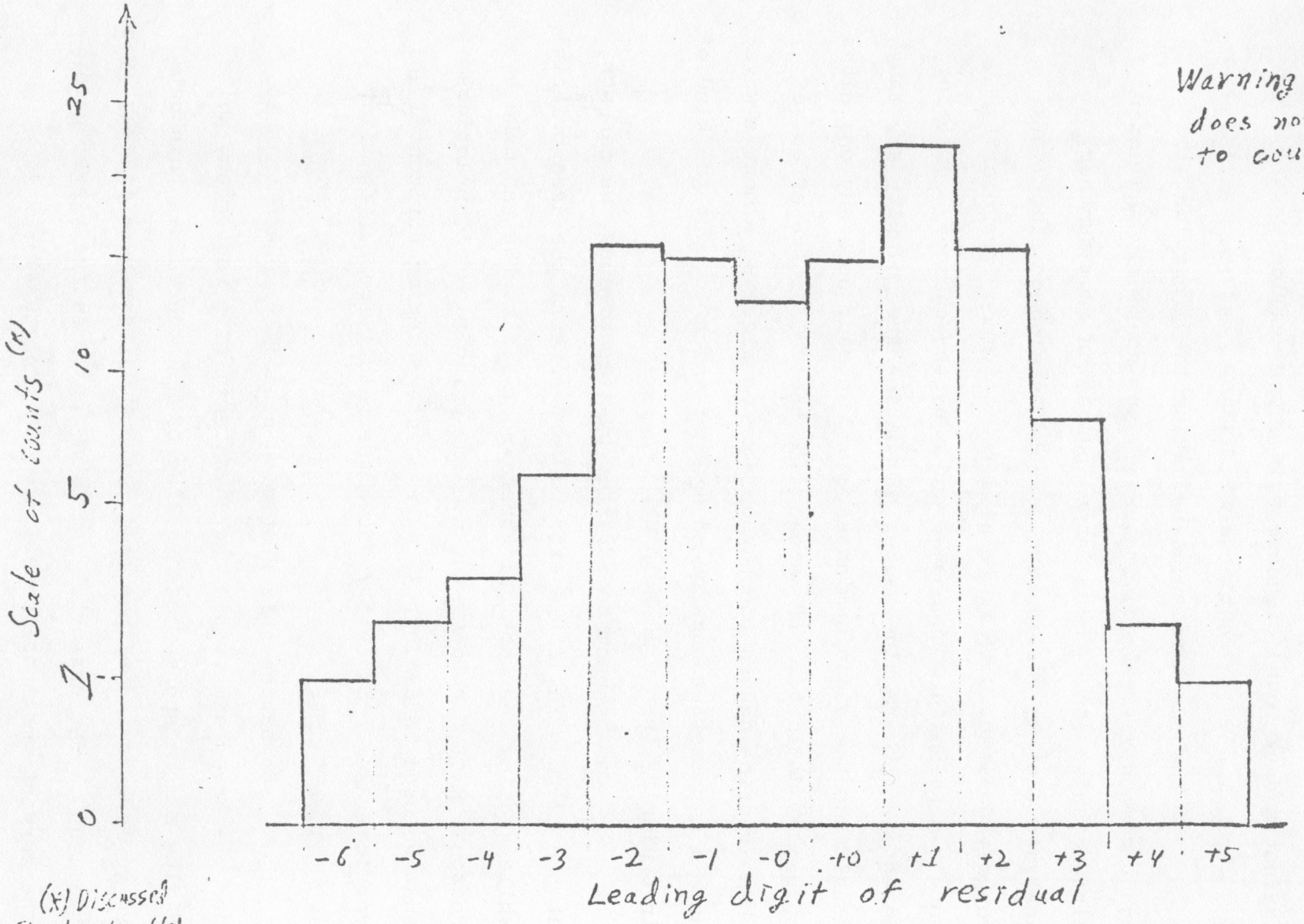
- the irregularities of outline now seem to be of more nearly equal size all across the distribution.
- the tails are now represented by larger, more noticeable blocks, so that we can see much more clearly what is -- or is not -- happening at or near them.

These observations seem to apply to the great majority of distributions of sets of values that arise in practice. They are part of the reason why square-root scales for counts are so useful in plotting distributions.

Some will wonder why we think it good that irregularities of outline have become more uniform in size. Is it not true, they will ask, that the search for revealing irregularity has been one of our main tasks? Are we not covering up just the sort of thing which we should be looking for? It is true that we seek revealing irregularity, but our problem is often -- indeed mainly -- a search for

exhibit 6 of chapter 25 -- Nebraska residuals

The by-now-familiar Example: Rootogram



(*) Discussed in text alick. Plot 10

revealing irregularity
against a background of unremitting, unrevealing,
routine irregularity

What we tend to do by making the irregularities that occur commonly (that is, frequently and routinely and under circumstances where there is no indication from further analysis that we should have detected a clue or clues) of roughly constant size is to give the uncommon irregularity both a more uniform chance -- and a better chance -- of being detected. The routine causes of routine irregularity are many -- usually no one cause dominates, nor, indeed, does any one class of causes dominate. If we can make the collective effect of these routine causes roughly uniform in size, we almost always have a much better background against which to spot the kind of irregularity that will help us.

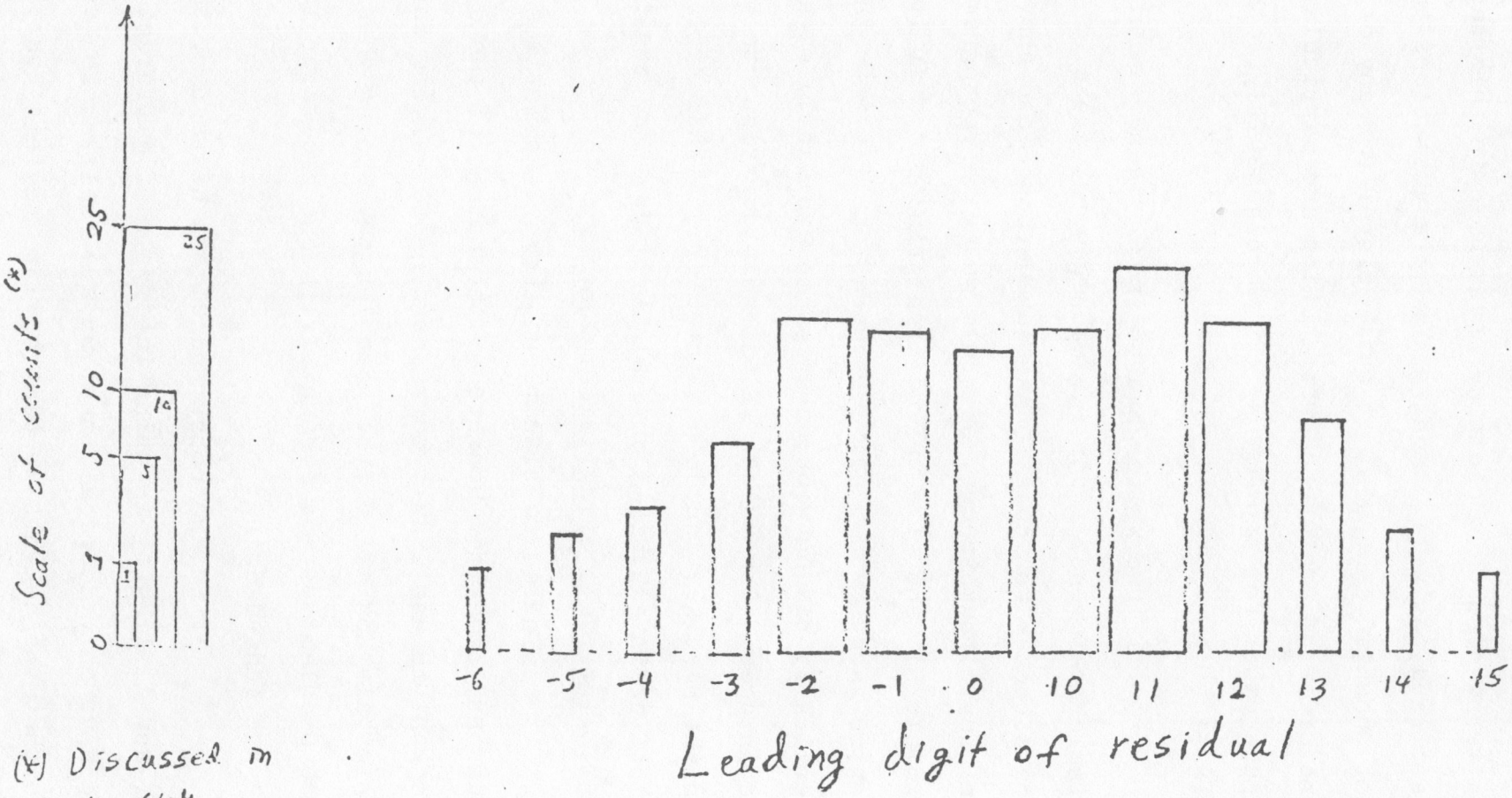
Coming back to our step-by-step advance, we can now, at the price of a little extra effort restore the area principle while keeping the vertical square-root scale. We can make a rootobar plot according to the same principles and practices that we used in exhibit 3, merely interchanging horizontal and vertical axes. Exhibit 7 shows the result for our familiar 120 values.

Comparing this plot with exhibit 6 we see that it has preserved almost all the advantages of a rootogram, while giving a more appropriate feeling of how the values thin out as we move toward either tail. (What matters here is exactly a "feeling". The essential question is: Does the viewer gain the correct emotional reaction to this picture?).

Different analysts will have different feelings about relative merits of different choices among the six forms of

exhibit 7 of chapter 25 -- Nebraska residuals

The by-now-familiar example: Conventional rootobar



(*) Discussed in text after Plot 10

patterns-of-counts plots we have shown in this section and the last (and, as well, among a few more variations that we have not shown). Indeed it is quite proper for any of us to have one order of preference when making one particular application and another for another -- or, more generally, others for others.

If the writer had to specify general rules fixing a single preference order for each general class of situation, he would choose to insist on a square-root scale above all else. In many circumstances, he would choose a rootogram (as in exhibit 6), and give up the area principle. If it was important not to shock rather naive readers familiar with this principle, vertically rising rootobars (as in exhibit 7) might have great advantages. Otherwise, if the considerations to be discussed in the next section do not interfere, either type of rootobar plot (as in exhibit 3 <u>or</u> exhibit 7) could be considered.

25D. <u>Intervals of different lengths</u>

There are situations where -- sometimes for extraneous reasons, sometimes because of the nature of the distribution, we want to use counts corresponding to varying intervals -- to intervals not all of the same length. Doing this raises some new problems -- and eventually throws a clearer light on what we are indeed doing.

If we are concerned with a histogram, there is little if any difficulty in using intervals of different lengths. We have only to recall that we are really using an area scale, so that the heights we plot have to be in a constant proportion to the ratios:

$$\frac{\text{number of values counted}}{\text{length of interval considered}}$$

Exhibit 8 shows how this works for our continuing example. In this plot we have used intervals for positive values just half the width of those used before, while continuing to use the old familiar intervals for negative values.

Except for an increased irregularity from rectangle to rectangle where the bases have been halved, the general impression given by this plot resembles reasonably closely that given by exhibit 5. This is clearly as we would hope it to be, though the increased irregularity can be the source of some difficulty.

Exhibit 9 shows the details and the almost-trivial arithmetic that is required.

The calculations necessary to make this plot have caused us to label the vertical scale differently than in exhibit 5 (and exhibits 4, 6, and 7). We have been more careful and correct by saying that the scale is really a scale of "counts per unit base" instead of a scale of "counts", and by showing the length of the unit base to which the counts apply.

We probably should have done just this in all the earlier plots.

Making a bar chart for intervals of different length raises no issues of arithmetic, but can raise issues of policy. If, as we are now convinced, we are trying to plot a rate -- the rate at which we encounter observations as we move along the scale, to speak of "counts per unit base" essentially admits

exhibit 8 of chapter 25 -- Nebraska residuals

The by-now-familiar example: Histogram with intervals of varying length

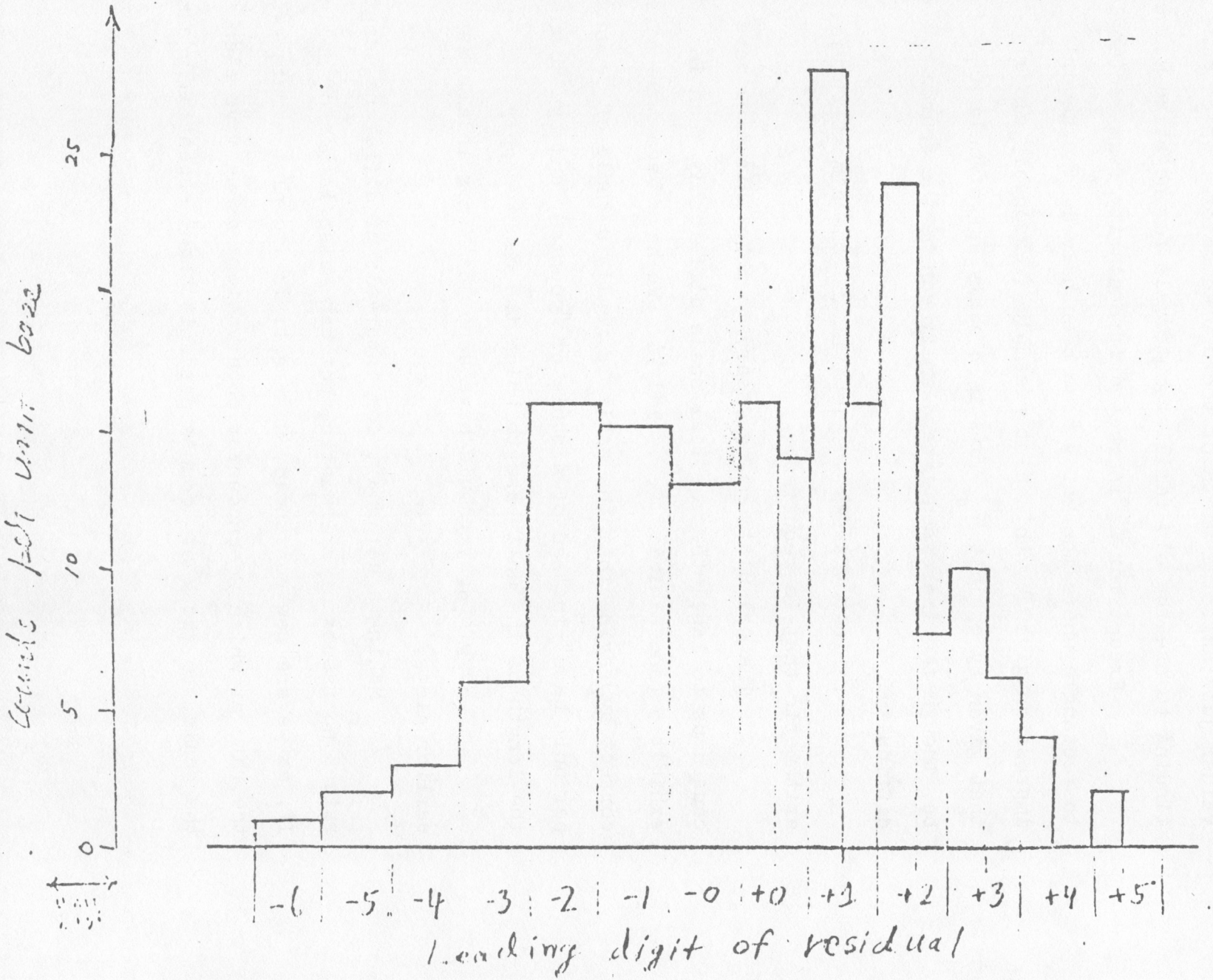


exhibit 9 of chapter 25 -- Nebraska residuals

Data and arithmetic for the same old example when positive intervals are of half width. (Counts can be checked from exhibit 1.)

Leading digit	Next digits	Count of values	Counts / length	(*)
+5	0 to 4	1	2	1.4
+4	5 to 9	0	0	0.0
+4	0 to 4	2	4	2.0
+3	5 to 9	3	6	2.4
+3	0 to 4	5	10	3.2
+2	5 to 9	4	8	2.8
+2	0 to 4	12	24	4.9
+1	5 to 9	8	16	4.0
+1	0 to 4	14	28	5.3
+0	5 to 9	7	14	3.8
+0	0 to 4	8	16	4.0
-0	all	13	13	3.6
-1	all	15	15	3.9
-2	all	16	16	4.0
-3	all	6	6	2.4
-4	all	3	3	1.7
-5	all	2	2	1.4
-6	all	1	1	1.0

(*) To be revealed in due course.

that this IS our interest when we use histograms -- we are supposed to copy the procedure for histograms as closely as we can. In particular, we can use the "counts/length" column of exhibit 9 to set the height of a bar chart presentation of the same example. Exhibit 10 does just this.

What of rootograms? Again nothing new, once we admit that the vertical scale is:

THE SQUARE ROOT of COUNTS/LENGTH

The last column of exhibit 9, marked (*), contains the square roots of the values in the previous column, all ready for use. Exhibit 11 shows the results of using them.

When we look carefully at the 2x2 array of plots of which this is the SE corner, namely:

	Equal intervals	Unequal intervals
Histogram:	exhibit 5	exhibit 8
Rootogram:	exhibit 6	exhibit 11

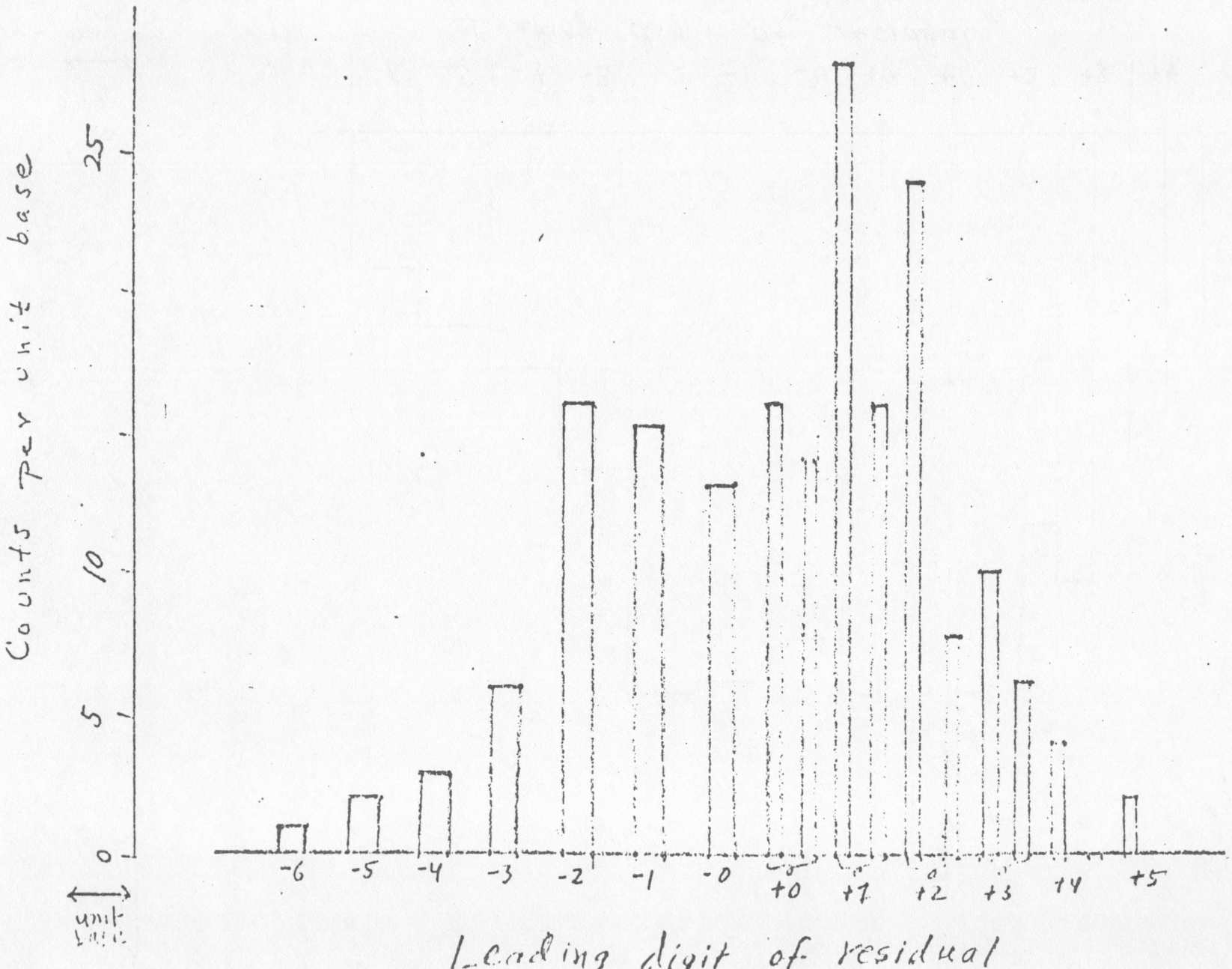
we realize an important fact:

- The effect on the rootogram of introducing unequal intervals is markedly less than the corresponding effect on the histogram.

Intervals of different widths lead, in all the techniques we are here considering, to different amounts of irregularity. For the histogram, this difference is combined with already existing differences in size of irregularity

exhibit 10 of chapter 25 -- Nebraska residuals

The by-now-familiar example: Bar chart for unequal intervals



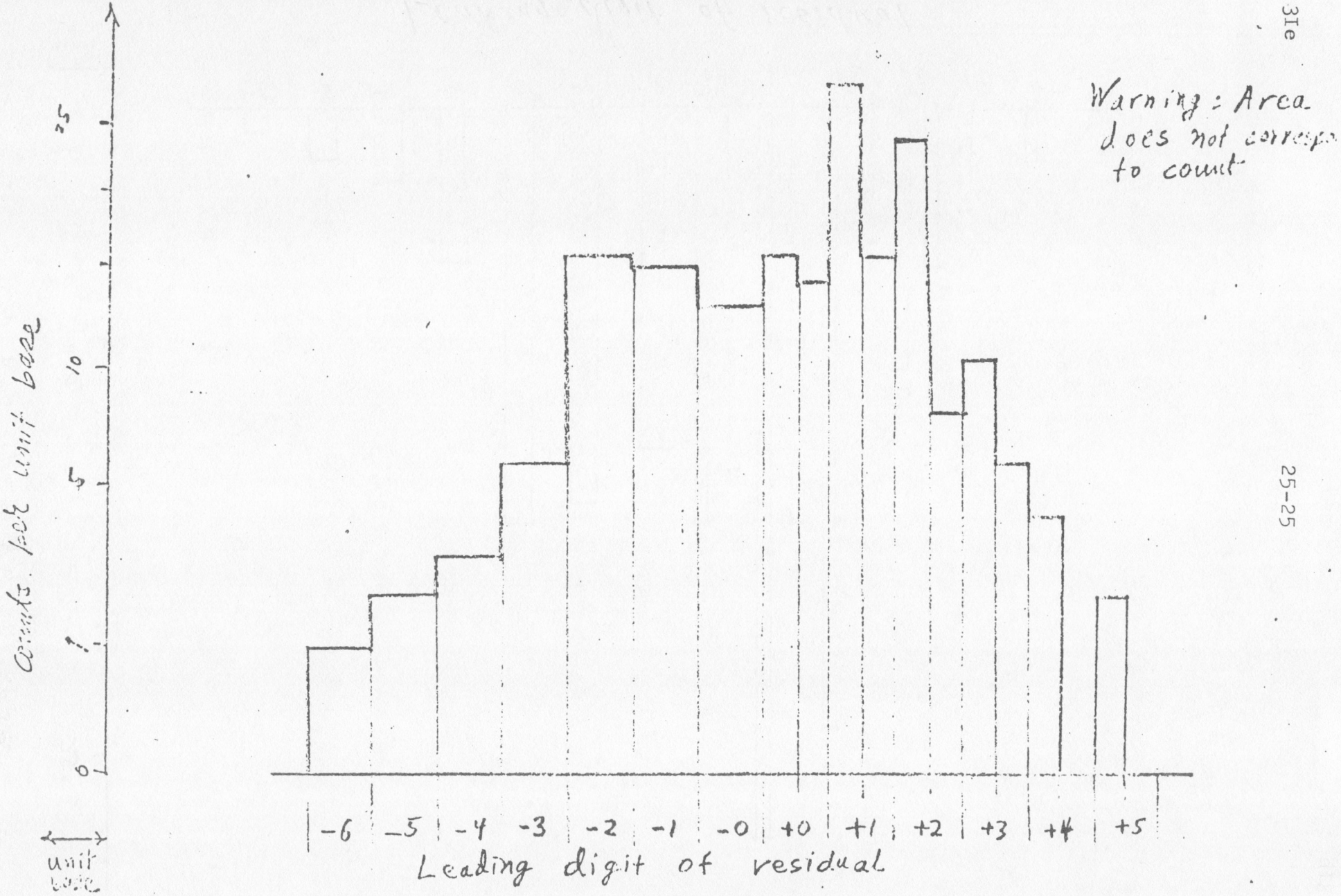


exhibit 11 of chapter 25 -- Nebraska residuals

The by-now-familiar example: Rootogram for unequal intervals

(larger where values are more frequent, smaller where they are not). Each source of difference in size tends to make the effects of the other worse. For the rootogram, which shows little change in size of irregularity when intervals are equal, the effects of inequality of interval width are less violent and less confusing. Again the rootogram proves more successful.

With both the square root and the counts/length ideas cleared up, a rootobar plot offers no difficulty. Exhibit 12 shows the result. This example displays our first sight of a general difficulty with rootobar plots. The total length of vertical lines in this picture is so great -- in comparison with the horizontal extent -- as to make the picture "busy" and somewhat confusing. After all, a rootobar plot requires about twice as many inches of vertical line as does the corresponding rootogram. We ought not to be surprised to find that we cannot afford to use as many intervals in a rootobar plot of given width and vertical scale as we can the corresponding rootogram.

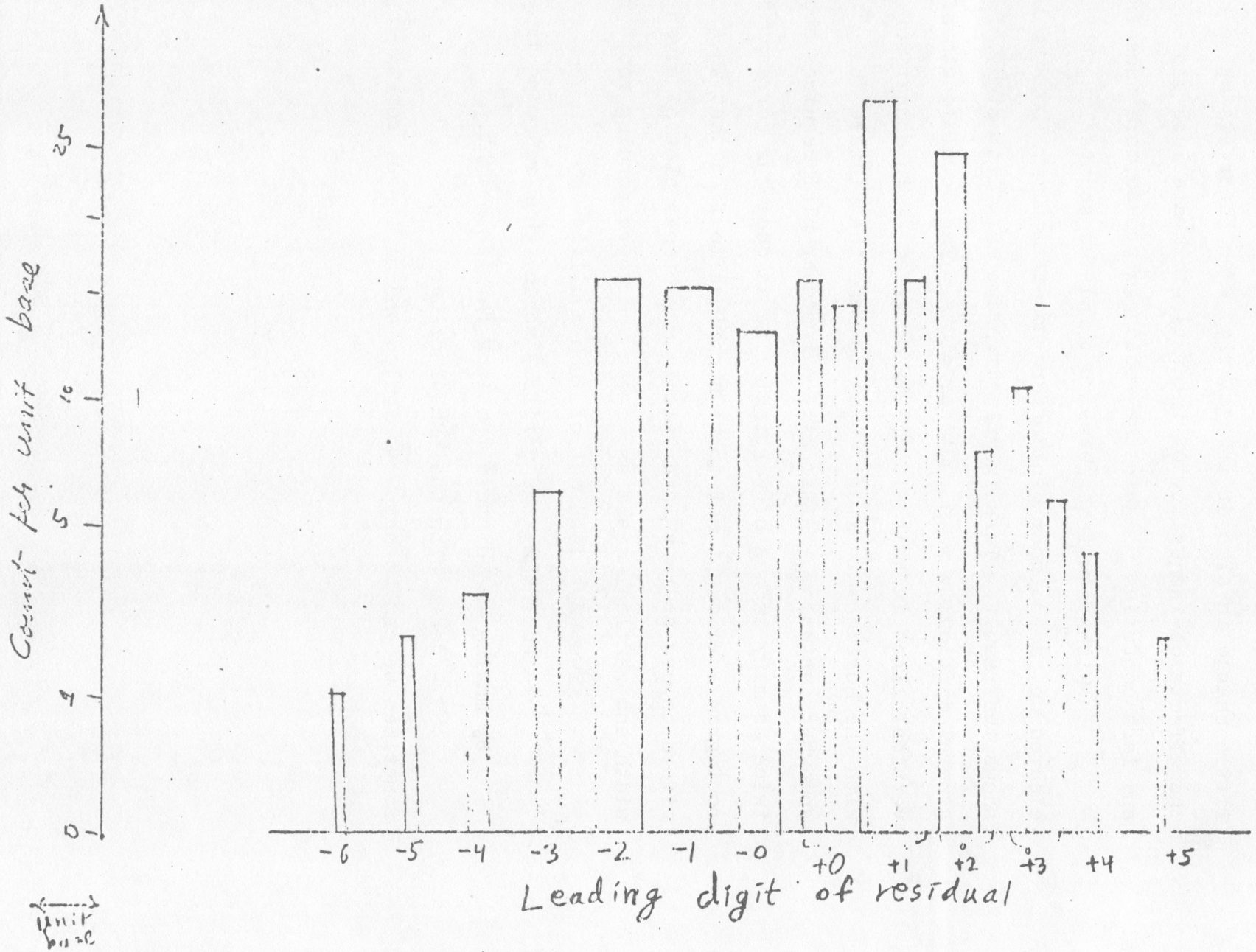
Two ways lead out of such a difficulty:

- shrinking the vertical scale, thus reducing impact by making the picture too long for its height.
- decreasing the number of bars.

Sometimes we can be happy to follow one of these, sometimes not.

exhibit 12 of chapter 25 -- Nebraska residuals

The by-now familiar example: Rootobar plot for unequal intervals



*****

25H. How far have we come?

This chapter has developed a variety of pictorial approaches to the display of information about distributions when there is nothing with which we are to compare:

- histograms, where heights are proportional to counts per unit base, and bases are what they are.
- rootograms, where heights are proportional to $\sqrt{\text{counts per unit base}}$, and bases are what they are.
- rootobars, where both heights and bases are proportional to $\sqrt{\text{counts per unit base}}$.

Three partially conflicting forces have to be resolved, often differently in different cases. They are:

- the usually overriding advantages of a square-root-of-counts-per-unit-length scale.
- the desirability of maintaining area proportional to count.
- the tendency of rootobar plots to become busy and confusing (and to involve more drafting effort).

If it were not for the third of these, we might always use rootobar plots. If it were not for the second, we might always use rootograms.

Indeed, if it were not for the first, we might always use histograms. In practice, situations that really indicate histograms are not too common -- the choice is usually between a rootobar plot and a rootogram. (Every analyst has to develop his own balance here.)

*****

25K. Xmas-tree patterns (background)

In section 25B we went from stem-and-leaf displays to various forms of "bars-flying-from-a-mast", the latter being a step on the way to the usual bar plots. We could have gone to more symmetric displays along the way, but did not because such displays do not seem specially useful. They are used, however, and are thus worth passing attention.

Exhibit 13 shows what happens when we try to make the stem-and-leaf pattern more symmetrical. Many displays in which left-to-right spread is symmetric about a vertical axis, of which this table illustrates only one kind, are likely to be called "Xmas trees" or "tree diagrams".

One important comment about the difference in character between exhibit 1 and exhibit 13 is in order. Exhibit 1 is a worksheet -- possibly polished up a trifle -- that also serves as a graphic display. Exhibit 13 is a picture that could not be a worksheet -- it could not have been easily started until all 120 values are at hand, since neither end of each line is fixed until we know how many entries go on that line.

Exhibit 14 sets out, in terms of bars of constant width, the same overall impressions conveyed by exhibit 13. Its pros and cons, in comparison with that table are mostly associated with choices of the ratio of width to height, choices that can clearly be made more freely and flexibly in a plot. Demographers like to use somewhat similar sorts of plot to show age distributions for the populations of various countries.

There is, of course, no reason why we should not use rootobars in a Xmas-tree display. Exhibit 15 does just this.

exhibit 13 of chapter 25 -- Nebraska residuals

Symmetrical (Xmas-tree) version of the stem-and-leaf array of exhibit 4 from chapter 4
(The ... mark, from top to bottom, the heights corresponding to
+5, 0, and -5, 0 respectively)

```
                    4
                  . . .
                   2 2
              1 1 8 3 3 9 2 6
          7 0 2 4 3 1 4 4 8 6 3 3 8 4 2 0
      2 6 3 1 3 3 3 6 7 9 4 9 7 9 2 2 5 0 0 1 4 3
          1 8 1 4 1 3 6 2 4 2 6 7 7 6 8
                    . . .
            1 6 7 0 4 8 4 4 3 7 4 3 5
          2 6 6 1 2 8 3 0 6 5 7 6 6 1 6
        9 3 0 0 1 6 0 7 0 7 9 1 6 4 5 6
                4 5 2 2 1 2
                  2 0 3
                   1 7
                  . . .
                    1
```

exhibit 14 of chapter 25 -- Nebraska residuals

Same distribution, summarized by bars of constant width

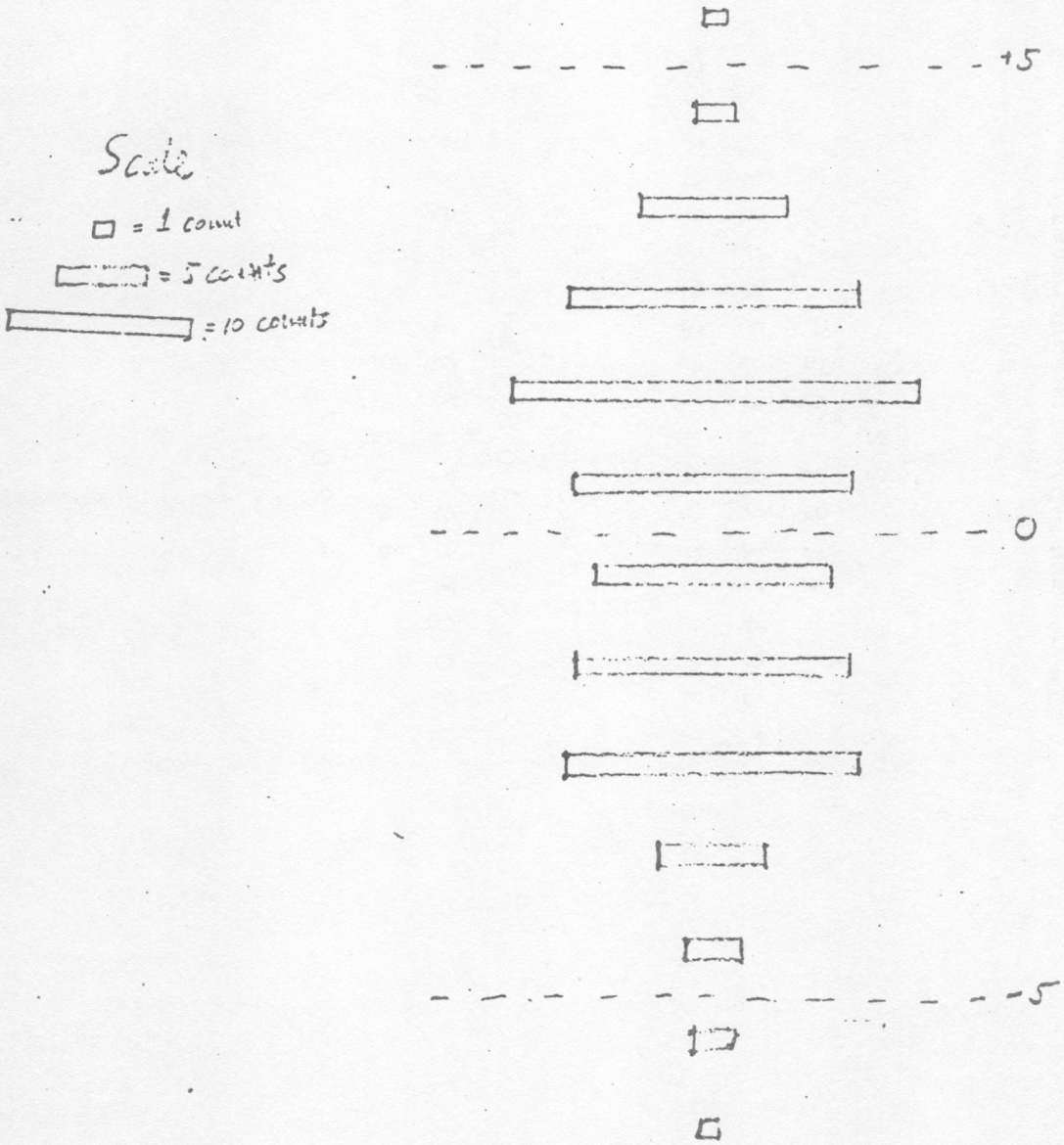
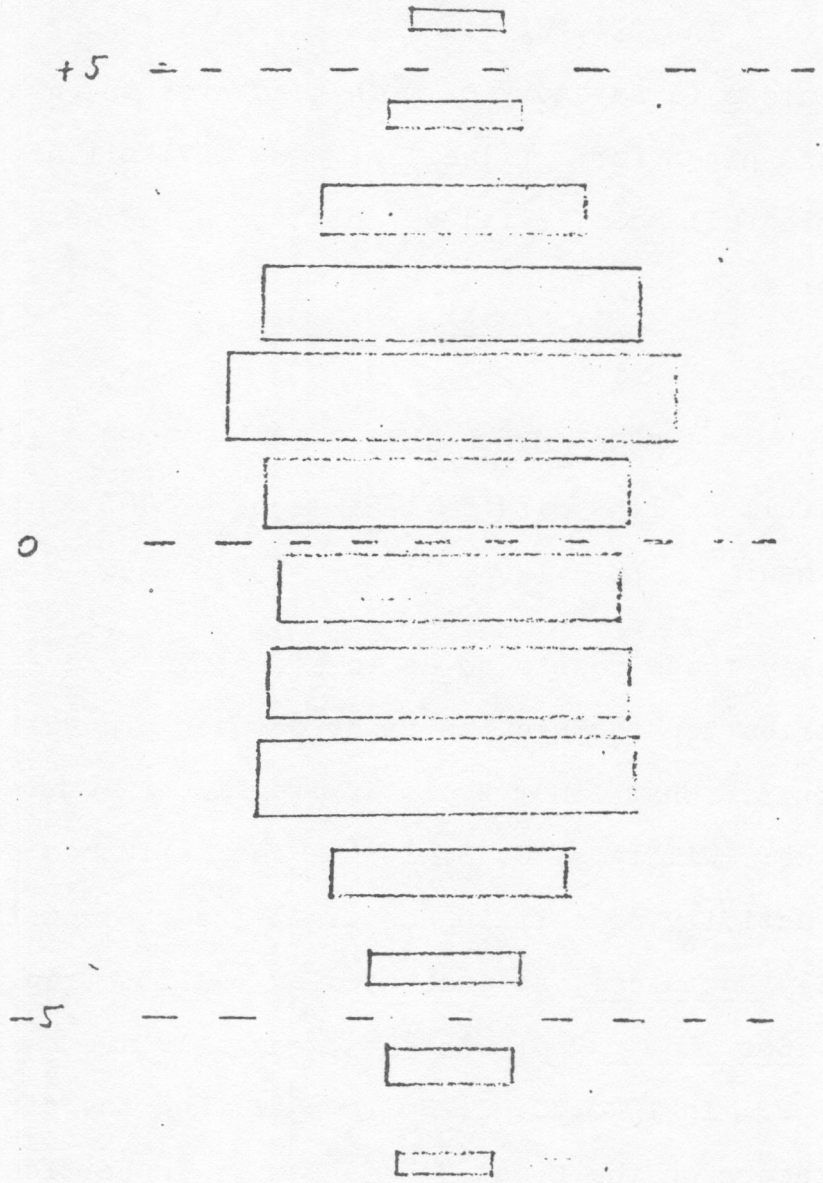
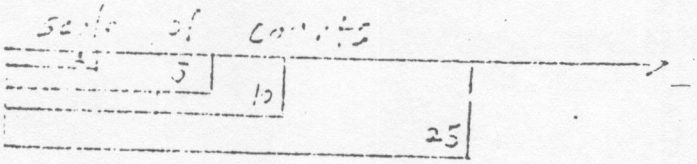


exhibit 15 of chapter 25 -- Nebraska residuals

The same data represented by rootobars
(Bars each of whose dimensions is proportional
to the square root of the corresponding count)





*****

25P. Problems for chapter 25

1. In 1936, Williams gave the following total numbers of Noctuidae (a category of moth) captured in each of eight periods of the night for: (i) all nights, (ii) all nights in the week before full moon, (iii) all nights in the weeks after full moon:

period:	1	2	3	4	5	6	7	8
All:	599	826	1813	1454	1759	1429	955	309
pre-full:	115	174	175	362	467	380	230	52
pre-new:	75	136	150	234	190	152	70	35

Display these counts so as to effectively reveal any change in distribution through the night between pre-full and pre-new weeks. Discuss. What other types of display might have been useful? Source: Williams, C. B. 1936. The influence of moonlight on the activity of certain nocturnal insects, particularly of the family Noctuidae, as indicated by a light trap. Phil. Trans. Roy. Soc. Lond. 226: 357-389, specially his table VII at page 374.

2. In 1940, C. I. Smith published the first of a series of papers on the Great Bahama Bank. In particular, he gave the number of days in each month during 1939 for which the wind at Nassau came from specified directions. His counts are given as exhibit 16. Omitting days of calm, plot the counts as revealingly as you can. Why did you select this way of presenting the data?

exhibit 16 of chapter 25

Wind Direction at Nassau, Bahamas Month-by-month During 1939

Month	Wind Direction								(Calm)
	N	NE	E	SE	S	SW	W	NW	
Jan	3	6	5	7	1	1	0	2	(6)
Feb	1	7	4	12	1	2	0	0	(1)
Mar	1	9	7	10	2	1	0	0	(1)
Apr	1	5	5	11	3	3	1	1	(0)
May	3	3	6	10	3	5	0	0	(1)
June	0	4	8	10	2	1	1	1	(3)
July	0	0	5	14	5	5	0	0	(2)
Aug	0	3	7	12	3	4	0	0	(2)
Sept	1	7	2	7	2	1	0	0	(10)
Oct	0	21	4	1	2	0	0	0	(3)
Nov	6	12	3	3	1	0	1	4	(1)
Dec	5	5	5	5	2	2	0	7	(0)

From: C. I. Smith 1951. The great Bahama bank. I. General hydrographical and chemical features. J. Marine Res. 3: 147-170, specially his table III, at page 155.

3. Brambell and Rowlands give counts of the numbers of certain rodents -- bank voles (Evotomys Glareolus) -- according to litter size as estimated in three ways: by the number of corpora lutea, by the number of placental sites, and by the number of embryos found (both in their work and by earlier workers), namely:

No:	1	2	3	4	5	6	7	8	more	(Tot)
Corp. Lut.	1	8	43	108	82	22	7	2	(*)	(273)
Pla. sites	-	1	11	20	13	11	2			(58)
Embryos	1	1	16	27	21	4	(own work)			(70)
Embryos	1	4	29	35	14	5	(earlier)			

(*) 2 at 9, 1 at 11, 1 at 12.

Plot the distributions in a revealing way.

Source: Brambell, F. W. Rogers and I. W. Rowlands 1936. Reproduction of the bank vole (Evotomys Glareolus, Schreber. I - The oestrous cycle of the female. Phil. Trans. Roy. Soc. London 226: 71-98, specially their Table IV on page 79.

4. Wake (1957) gave frequencies of sentence lengths of two editions of Aristotle's Categories, as repeated in exhibit 17. Make comparative rootograms of these distributions. Comment on the results.

5. In 1930, T. V. Pearce reported the number of women aborting at specified term in weeks. Exhibit 18 has the numbers. Make a revealing plot.

exhibit 17 of chapter 25

Distributions of Sentence Lengths for Two Editions of Aristotle's Categories

Sentence length (words)	Number of sentences	
	Oxford	Loeb
4	2	3
5	7	12
6	15	15
7	21	20
8-10*	41	51
10*-12	35	49
13-15	52	51
16-19	48	42
20-25	55	41
26-31	23	26
32-39	11	12
40-50	1	1
51-63	1	1
64-79	-	2

B) SOURCE: William C. Wake 1957. Sentence length distributions of Greek authors. J. Royal Statist. Soc. Ser. A, 120: 331-346, Table 2 at page 336.

exhibit 18 of chapter 25

Number of Women Aborting at Specified Term

A) DATA

Term in weeks	Count
4	3
5	7
6	10
7	13
8	14
9	29
10	22
11	21
12	18
13	28
14	16
15	19
16	10
17	13
18	14
19	8
20	4
21	2
22	10
23	4
24	4
25	3
26	4
27	6
26	1

B) SOURCE

T. V. Pearce 1930. An unusual frequency distribution - The term of abortion. Biometrika 22: 250-252. Used by: M. G. Kendall and A. Stuart 19--. The advanced theory of statistics (3rd Ed.) 1: 27.

6. In 1914, Karl Pearson reported the lengths of 2500 trypanosomes from Glossina morsitans. Exhibit **19** has the data. Plot.

7. In 1930, T. L. Woo reported the numbers of old Egyptian skulls with specified lengths of the left occipital bone in millimeters. Exhibit **20** has the data. Plot it sensibly.

8. In 1948 E. E. A. Archibald published the counts of numbers of plants per quadrat set out in exhibit **21**. Make a rootobar plot.

exhibit 19 of chapter 25

Lengths of 2500 Trypanosomes

A) DATA

Length (microns)	Count
15	7
16	31
17	148
18	230
19	326
20	252
21	237
22	184
23	143
24	115
25	130
26	110
27	127
28	133
29	113
30	96
31	54
32	44
33	11
34	7
35	2

B) SOURCE

K. Pearson 1914. On the probability that two independent distributions of frequency are really samples of the same population, with special reference to recent work on the identity of trypanosome strains. Biometrika 10: 87-143. See page 112. Used by: M. G. Kendall and A. Stuart 19--. The advanced theory of statistics (3rd Ed.) 1: 11.

exhibit 20 of chapter 25

Lengths of the Left Occipital Bone for Egyptian Skulls

A) DATA

Length (in mm*)	Count
85h	12
87h	12
89h	32
91h	48
93h	79
95h	116
97h	104
99h	126
101h	123
103h	74
105h	68
107h	36
109h	18
111h	7
113h	4
115h	4
117h	-
119h	1

B) SOURCE

T. L. Woo 1930. On the asymmetry of the human skull. Biometrika 22: 324-352. Table XXIX on page 351. Used by: M. G. Kendall and A. Stuart 19--. The advanced theory of statistics (3rd Ed.) 1: 25.

* Centers of intervals; "h" means "and a half".

exhibit 21 of chapter 25

Numbers of Plants of Salicornia Stricta per Quadrat

A) DATA

Plants	Count
0	4
1	3
2	8
3	13
4	11
5	9
6	8
7	10
8	3
9	3
10	8
11	3
12	4
13	4
14	-
15	3
16	-
17	-
18	1

B) SOURCE

E. E. A. Archibald 1948. Plant populations I. A new application of Neyman's contagious distribution. Annals of Botany 12: 221-235. Table - on page 228.